AWARD NUMBER: W81XWH-16-1-0194

TITLE: Molecular and Cellular Determinants of Malignant Transformation in Pulmonary Premalignancy

PRINCIPAL INVESTIGATOR: Kostyantyn Krysan

CONTRACTING ORGANIZATION: University of California, Los Angeles
Los Angeles, CA 90095

REPORT DATE: July 2017

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
July 2017	Annual	1 Jul 2016 - 30 Jun 2017

4. TITLE AND SUBTITLE
Molecular and Cellular Determinants of Malignant Transformation in Pulmonary Premalignancy

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-16-1-0194

5c. PROGRAM ELEMENT NUMBER

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

6. AUTHOR(S)
Kostyantyn Krysan

E-Mail: KKrysan@mednet.ucla.edu

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of California, Los Angeles
10889 Wilshire Blvd, Suite 700
Box 951406, LOS ANGELES CA 90095-1406

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT
During the first funding period we completed Major Tasks 1 and 2. The areas of interest were identified in 41 lung cancer patient and isolated by LCM. Genomic DNA was isolated from these areas and whole exome sequencing was performed. The data has been analyzed and the progression-associated mutations, as well as premalignant- and malignant-specific mutations, were identified. The mutational data was analyzed in the pathway context. Based on the mutational analysis, neoantigens were identified.

15. SUBJECT TERMS
Lung cancer, premalignancy, progression, whole exome sequencing, driver mutations, neoantigens.

16. SECURITY CLASSIFICATION OF:

a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	Unclassified	11	USAMRMC

19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. Introduction

Lung cancer is the leading cause of cancer death in the U.S. and throughout the world with adenocarcinoma being the leading subtype in the US, Japan and other countries. One of the major driving forces of carcinogenesis is somatic mutagenesis. Over 75% of lung cancers bear driver mutations that are causally implicated in cancer development, while the remainder of lung cancers does not bear mutations in known oncogenes or tumor suppressors. Distal airways of many lung cancer patients and subjects at risk for developing lung cancer often contain small focal proliferative lesions designated atypical adenomatous hyperplasia (AAH). Current studies suggest that AAH may be a precursor of adenocarcinoma *in situ* (AIS) and, subsequently, to invasive pulmonary adenocarcinoma (ADC). Factors that determine the fate of a premalignant lesion, i.e. whether it will progress to cancer or recede, remain enigmatic. Early attempts to evaluate somatic mutations in premalignant pulmonary lesions revealed mutations in known driver genes, such as *KRAS*, *EGFR* and *TP53*. Furthermore, clonal analysis demonstrated identical monoclonal patterns in AIS and AAH adjacent to it, strengthening the notion that AAH is a preneoplastic lesion rather than reactive hyperplasia. A recent study utilizing targeted sequencing of AAH lesions and related tumors identified mutations in other cancer-related genes as well as clonality between premalignant lesions and cancer. This study also highlighted the importance of the mutational landscape variations in progression from premalignancy to cancer, however, the genomic and microenvironmental determinants of progression have not yet been elucidated.

2. Keywords

Lung cancer, premalignancy, progression, driver mutations, neoantigens, whole exome sequencing (WES).

3. Accomplishments

- **What were the major goals of the project?**

Specific Aim 1(specified in proposal)	Timeline	Site 1	Status after Year 1
Major Task 1	Months		
Subtask 1: Review the slides to identify the areas of interest for LCM and IHC	1-3	Dr. Wallace	Completed
Subtask 2: To isolate areas of interest by LCM	2-4	Dr. Krysan	Completed
Subtask 3: To isolate genomic DNA and perform quality control	5	Dr. Krysan	Completed
Milestone(s) Achieved			Completed
Local IRB/IACUC Approval: Active, IRB#10-001096-CR-00005		Dr. Krysan	Completed
Milestone Achieved: HRPO/ACURO Approval			Completed
Major Task 2			
Subtask 1: To construct sequencing libraries and perform exome enrichment (50 cases)	6-8	Dr. Krysan	Completed
Subtask 2: To perform next generation sequencing	9-11	Sequencing Core facility	Completed
Subtask 3: To perform data analysis and identify progression-associated mutations	12-14	Drs. Krysan and Tran	Completed
Milestone(s) Achieved:			All
Specific Aim 2			
Major Task 3			
Subtask 1: To perform multi-color IHC, slide scanning and image analysis	15-20	TPCL, Dr. Wallace	Ongoing
Subtask 2: To relate the expression of immune regulators to the mutational landscapes of the tissues	21-24	Drs. Tran and Krysan	Ongoing
Milestone(s) Achieved:			

- **What was accomplished under these goals?**

During the first funding period we fulfilled Major Tasks 1 and 2. We performed WES and identified putative neoantigens in 89 AAH, 15 AIS, and 55 ADC lesions from 41 lung cancer patients (**Table 1**). The cells of interest were dissected from the following regions of distal airways utilizing Laser Capture Microdissection (LCM): **a**) normal airway epithelial cells (1-3 regions per patient), **b**) 2-4 AAH lesions, **c**) 1-3 AIS (where present), and **d**) 1-3 regions of ADC, including multiple primary lung tumors (where present). Sequencing libraries were constructed followed by exome enrichment and WES was conducted with at least 2 x 10^{10} bases sequenced per exome, which has been frequently achieved in published WES studies.

Patient ID	AAH	AIS	ADC
P01	2	1	3
P02	2	0	1
P03	2	1	2
P04	3	0	3
P05	4	0	4
P06	3	0	1
P07	1	2	1
P08	3	0	1
P09	2	0	1
P10	2	1	2
P11	2	0	1
P12	3	0	4
P13	3	0	2
P14	2	1	1
P15	2	0	1
P16	2	0	1
P17	1	1	0
P18	4	0	1
P19	1	0	1
P20	3	0	1
P21	2	0	1
P22	2	0	1
P23	2	0	1
P24	2	0	3
P25	2	0	1
P26	2	1	1
P27	2	1	2
P28	2	0	1
P29	2	0	1
P30	2	1	2
P31	3	0	1
P32	2	2	0
P33	2	0	1
P34	1	1	1
P35	2	0	1
P36	2	1	1
P37	2	0	1
P38	2	0	1
P39	2	1	0
P40	2	0	1
P41	2	0	1

Table 1. Summary of regions sequenced in each patient.

The median number of mutations identified per individual region was 351. The median total of all non-synonymous (n.s.) somatic mutations in all regions sequenced per patient was 1323. The mutational load per patient did not increase significantly by the addition of more sequenced regions (Kruskal-Wallis rank sum test p = 0.20) (**Figure 1**).

To characterize heterogeneity among sequenced regions. we utilized the Jaccard index, which measures the similarity in n.s. somatic mutations between a pair of lesions, and is inversely proportional to the level of heterogeneity. We found that lesions obtained from within individual patients had significantly higher Jaccard indices and, thus, lower heterogeneity than lesions compared between patients (Kruskal-Wallis rank sum test p < 10-16) (**Figure 2A**). By further examining the heterogeneity between regions in individual patients, we found that their indices varied over a wide range (**Figure 2B**). With the exception of the first four patients (P01 — P04, **Figure 2B**), individual patients had higher indices (lower heterogeneity) among regions as compared to those from different patients. Thus, in this cohort, each patient most often demonstrated unique n.s. somatic mutations not shared among patients.

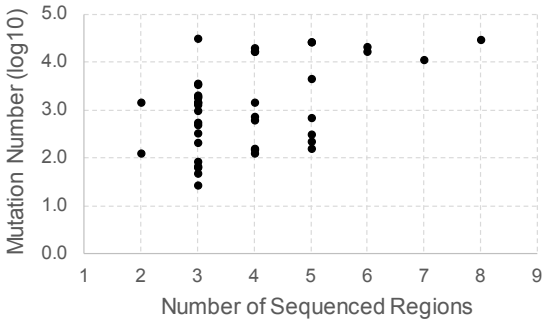

Figure 1. The number of lesions sequenced per patient does not significantly alter the mutational load.

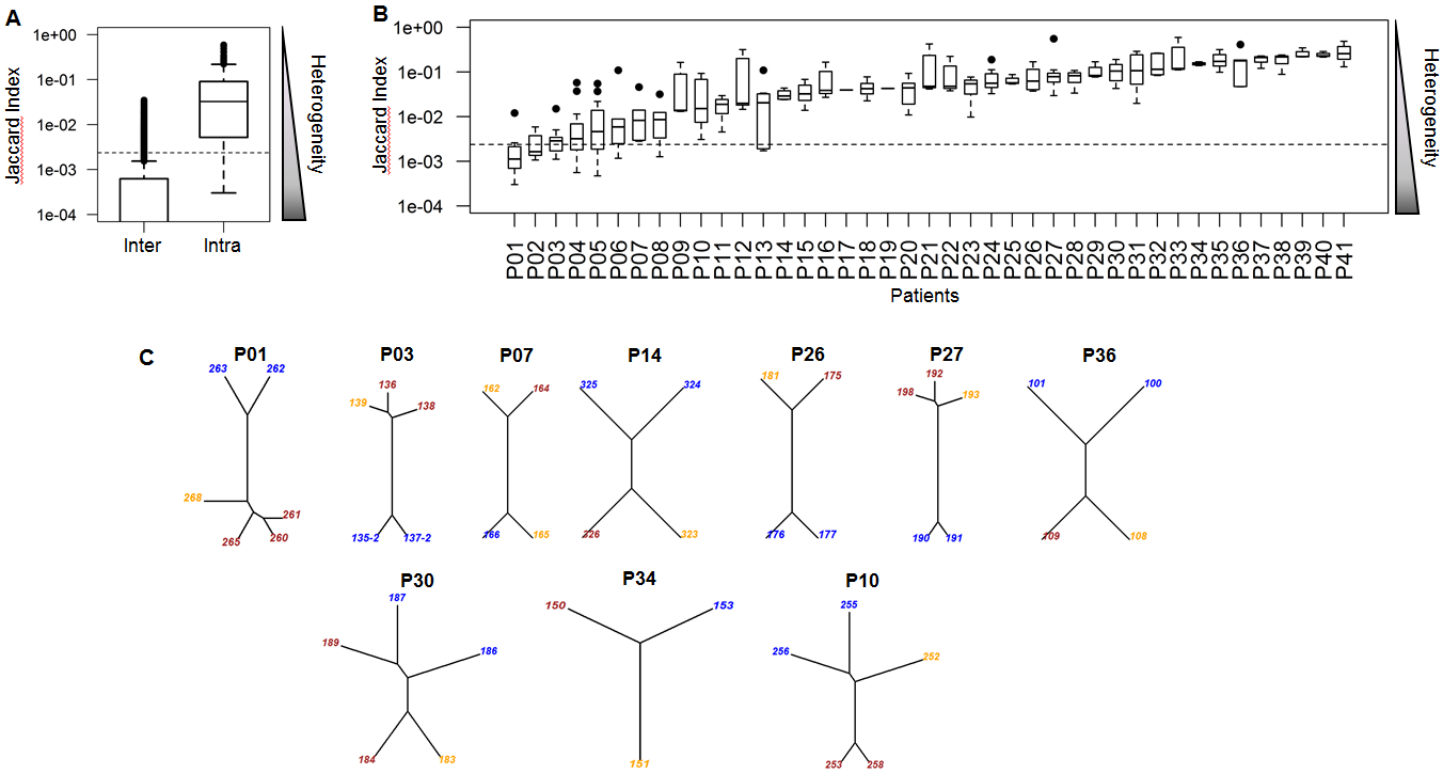

Figure 2. Intra- and inter-patient genetic heterogeneity of pulmonary lesions.
A) Distribution of Jaccard indices comparing n.s. somatic mutation heterogeneity between AAH lesions from the same (intra-) or different (inter-) patients. Inter-patient Jaccard indices have the zero-median, which cannot be displayed on the log-scale y-axis, therefore, the distribution starts at 54 percentiles. **B**) Distribution of intra-patient Jaccard indices in 41 individual patients. The subjects are displayed in the low-to-high order based on their median values. In **A** and **B**, the side triangles represent the heterogeneity levels inversely proportional to Jaccard indices, and the dashed line marks the 90 percentile level of inter-subject Jaccard index. **C**) Phylogenetic trees for 10 patients with AAH (blue), AIS (orange) and ADC (brown).

AAH lesions are pathologically classified as pulmonary premalignancy and AIS — as non-invasive malignancy, however, their somatic mutations have not been fully examined along with those in ADC to confirm if the analogous classification should be applied based on genomic profiles. AAH, AIS and ADC were all present in 10 out of 41 studied patients and their mutation profiles were compared to determine the analogy among

regions. Phylogenetic trees were constructed for all cases to illustrate the relationship between lesions (data for ten cases that had AAH, AIS and ADC are shown in **Figure 2C**). Interestingly, ADC shared common mutations with AIS, but not with AAH in most cases, except **a**) one case in which one of two primary ADCs in the patient was closely related to AAH lesions, while another ADC was clustered to AIS, **b**) one case in which ADC had its mutational landscape highly overlapped with that of AAH but not of AIS, and **c**) one case in which mutations in AAH and AIS lesions were closely related to each other, but not to ADC. Thus, in our further studies we categorized AIS as malignant lesions because in most cases their mutational landscapes closely resembled those for ADC.

To determine how n.s. somatic mutations affect tumor development in various stages, we first classified them into three different categories: **a**) premalignant mutations (PrMs) which were observed only in AAH lesions, **b**) progression-associated mutations (PAMs) which were located in both AAH and AIS/ADC lesions, and finally **c**) malignant-specific mutations (MSMs) which were only identified in AIS/ADC lesions. Recent studies, focused on intra-tumoral heterogeneity, have classified mutations as trunk (or clonal), branch and private (subclonal) mutations. Here, the classification is based on the type of the lesion where the mutations are located. Therefore, PAMs are comprised of trunk and branch mutations, while MSMs are composed of branch and private mutations. The median number of mutations identified per individual region was 351. The median total of all of all mutations in all regions sequenced per patient was 1323. However, the mutational load per patient did not increase significantly by the addition of more sequenced regions (Kruskal-Wallis rank sum test p = 0.20). Similarly, the percentage of PAMs varied over a wide range (0.2% to 44%), and was anti-correlated to the numbers of sequenced regions. The variation in PAM levels due to change in regions number was insignificant (Kruskal-Wallis test p = 0.24). The high variation of the PAM percentage reflected diversity in the mutational profiles among patients in our cohort.

Next, the association of neoantigens generated by PAMs with immune cell infiltration was investigated. Putative neoantigens were derived from n.s. somatic mutations to determine their association with apparent adaptive immune responses, as reflected by T cell infiltration and upregulation of checkpoints in premalignant and malignant tissues. Multiple algorithms were applied to predict binding affinity (IC_{50}) between mutant proteins and patient HLAs based on the Immune Epitope Database recommendations. Mutant peptides with predicted IC_{50} < 500 nM were considered neoantigens. In accordance with our mutation classification, the neoantigens were also categorized into three groups as premalignant (PrNs), progression-associated (PANs) and malignant-specific (MSNs) neoantigens. As expected, the total number of putative neoantigens per patient was highly correlated with the corresponding mutational load (Kendall's τ = 0.90). The distribution of neoantigen groups in 41 cases is summarized in **Figure 3A,** in which the cases are ordered based on the percentage of PANs. The percentage of PANs per patient varied from 0.2 to 40% with a median of 5%, while that of MSNs fluctuated from 5% to 92%, and 6% to 90% for PrNs. In addition to the patient level analysis, neoantigens were characterized in each specific region. For example, the percentages of PANs in the individual AAH lesions were similar to the associated cancer, whereas the percentage of PANs at the patient level demonstrated inter-region homogeneity. **Figure 3B** illustrates the variation in the percentage of PANs in individual AAHs for each case arranged based on the median observed level.

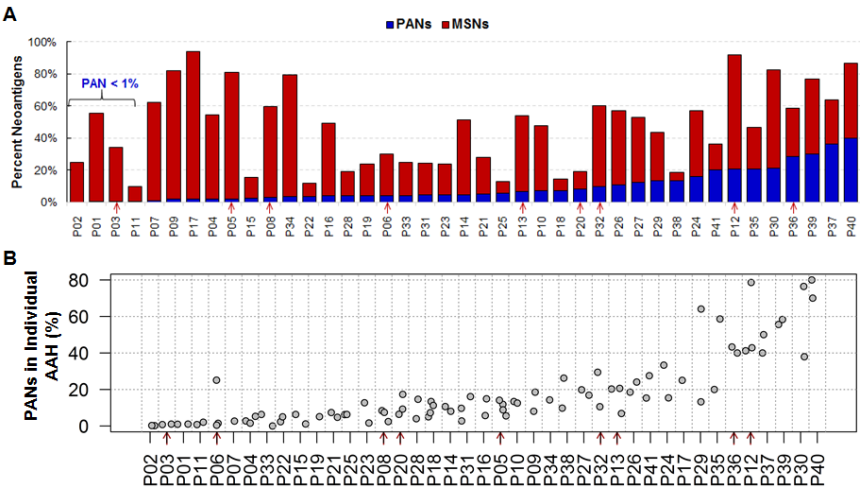

Figure 3. Neoantigens and the immune response in pulmonary premalignancy.
A) Distribution of PANs and MSMs in 41 study patients. The patients are displayed in the low-to-high order based on their percentages of PANs. Red arrows in **A** and in **B** indicate nine patients whose cellular immune response will be evaluated. **B**) Percentage of PANs in individual AAH regions from 41 patients. The subjects are displayed in the low-to-high order based on their median levels, and not in the same order as those in **A**.

Next, we evaluated the mutational status of 29 driver genes frequently mutated in lung ADC (identified by the TCGA study) in our cohort of patients. We found that these genes were frequently mutated in ADC, but rarely in AAH. These results suggest that the driver mutations were necessary for the progression from AAH to cancer. Comparing our data with TCGA LUAD (lung adenocarcinoma cohort), we found that the frequencies of *TP53*, *KRAS* and *KEAP1* mutations were lower in our data set, while those of growth factor receptors (such as *EGFR* and *ERBB2*) were higher. Although driver gene mutations are important for tumor development, they are absent in 24-36% lung ADC patients. In our cohort, n.s. mutations in the above mentioned 29 driver genes were absent in 51% of patients.

Mutated genes in ADC included those bearing both PAMs and MSMs; therefore, it was essential to determine the input of each of the gene groups in the pathway context. The ES of each gene group was calculated for all 1341 pathways. **Figure 4A** shows the recurrence rate of the top 27 pathways that are frequently deregulated by the genes bearing MSMs. These pathways were also affected by the PAM-bearing genes, but predominantly at a lower frequency than the MSM-bearing genes. The termination of O-glycan biosynthesis pathway was found to be affected by the PAM-bearing genes in 85% of patients. This glycoprotein sialylation pathway involves several mucin proteins, including *MUC4*, that have been found to bear PAMs in 90% of the patients in the current cohort. Among pathways de-regulated by MSM-bearing genes, the focal adhesion pathway was ranked highest based on the recurrence rate. In 30 of 41 patients, this pathway was affected by either PAM- (n = 1) or MSM- (n = 17) bearing genes or both (n = 12) (**Figure 4B**). These patients were divided into two groups: Group 1, in which pathways were affected by both PAM and MSM bearing genes, and Group 2, in which pathways were affected only by MSM bearing genes. We found that: **a**) the Group 1 patients had PAM-based ES higher than the MSM-based ES (Kruskal-Wallis p = 0.01), and **b**) in the Group 2 patients the MSM-based ES was generally higher than that in Group 1 (Kruskal-Wallis p = 0.0004). As the ES is proportional to the percentage of the mutated genes belonging to the pathway of interest, these results imply that if a high percentage of PAM-bearing genes is involved in the oncogenic pathway, few additional MSMs are required for its de-regulation (**Figure 4B**, Group 1). In contrast, other pathway activities are affected only by the MSM-bearing genes (**Figure 4B**, Group 2).

In the next step we comprehensively evaluated deregulation of all 1341 pathways based on PAM- and MSM-bearing genes to gain insight into tumor initiation and development. Deregulation status of all pathways based on two mutation groups (PAM or MSM) was tabulated as a one-and-zero binary matrix for all patients. The unsupervised hierarchical cluster analysis based on the pathway status identified three patient groups designated as high (H, n = 12), intermediate (I, n = 20), and low (L, n = 9) depending on the number of pathways deregulated by PAM- and MSM-bearing genes (**Figure 4C**). The intermediate group included the majority of study patients, in which

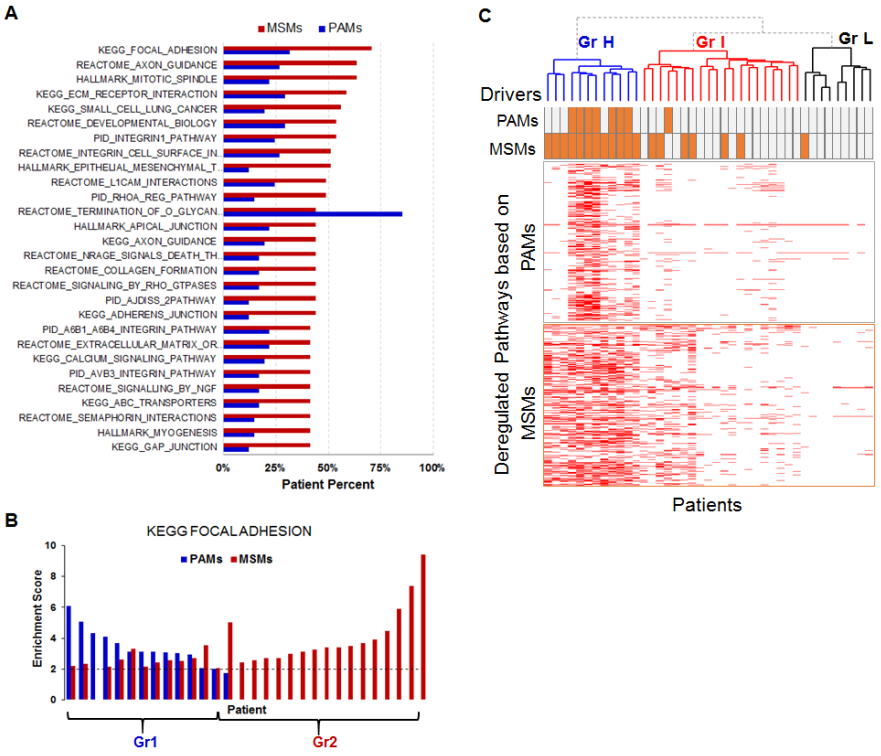

Figure 4. Analysis of pathway deregulation by PAM and MSM.
A) The top 27 pathways frequently affected by MSM- (red) and PAM- (blue) bearing genes. **B**) Enrichment score of MSM- (red) and PAM- (blue) bearing genes involved in the KEGG-based focal adhesion pathway, which is the most deregulated pathway by MSM-bearing genes, plotted for each patient. Patients having less than two pathway genes (i.e. ES = 0) mutated in ADC are not included in the analysis. The gray dashed line indicates the significant ES threshold (= 2) to determine if the genes involved in pathway were significantly enriched by the mutated genes in the specific group. Two patient groups were defined based on their PAM-based ES ≥ 2. **C**) Heatmap of the pathways affected (red) by PAM- (top) and MSM- (bottom) bearing genes. The mutations in the 29 drive genes observed in PAM and MSM are indicated by orange bars above the heatmap.

MSMs (but not PAMs) were the main source of pathway deregulation and were frequently found in the driver genes. Overall, it appears that in this group MSMs in the driver genes were essential for malignant progression. Group L, the smallest group, had infrequent pathway deregulation by either PAM- or MSM-bearing genes. Group H had the highest number of deregulated pathways among the three groups. The deregulated pathways in this group were frequently affected by both PAMs and MSMs, as well as driver genes (**Figure 4C**). Also, based on the deregulation pattern of the focal adhesion pathway, the majority of the patients in this group belong to Group 1 shown in **Figure 4B**. In Group H, *PIK3CA*, *PIK3R3* (catalytic and regulatory subunits of PI3K, respectively) and *PPP2R1A* genes (regulatory subunit of phosphatase PP2A negatively regulating AKT kinase) had high frequency of somatic mutations in AAH lesions. However, in all patients with PAMs in either or both *PIK3CA* and *PPP2R1A* genes, these mutations were detected in only one AAH lesion, and thus appeared as branch mutations. This suggests that the PI3K/AKT pathway deregulation required another pathway to be deregulated to induce malignant transformation. Similarly, higher numbers of deregulated pathways in group H suggest that deregulation of multiple non-critical pathways may synergize with those that are critical, leading to malignant transformation. Conversely, in group L there were very few pathways deregulated by both PAM- and MSM-bearing genes, suggesting that the transformation could be caused by events other than the somatic driver mutations that were not readily detectable by WES, such as gene rearrangements, copy number variation, epigenetic changes, deregulation of gene expression or alternative splicing. The fact that group L included both patients that only had AIS but no ADC, suggests that lesions in this group had a low invasive potential. The majority of study patients had driver genes and pathways affected by MSMs, suggesting that the mutation profiles of their AAH lesions were less complex compared to the associated ADC.

- **What opportunities for training and professional development has the project provided?**

These preliminary results were presented as the oral presentation at the American Association for Cancer Research 2017 annual meeting in Washington, DC.

- **How were the results disseminated to communities of interest?**

As these results are preliminary, they have not yet been disseminated. After the completion of the project, the WES data will be deposited to the publicly available data repository.

- **What do you plan to do during the next reporting period to accomplish the goals?**

During the next funding period we will fulfill the Major Task 3, including the multi-color IHC, slide scanning, image analysis and linking the expression of immune regulators to the mutational landscapes of the tissues

4. Impact.

- **What was the impact on the development of the principal discipline(s) of the project?**

In our preliminary studies we first demonstrated that heterogeneity between different AAH lesions from the same patient is significantly lower than between lesions from different patients. This leads to the occurrence of different PANs, which in turn will cause varied treatment responses and outcomes. On the other hand, low genomic complexity of pulmonary premalignancies raises hope that unleashing the immune response against them (as opposed to targeting the established tumors) may be a successful strategy for cancer prevention. These findings clearly demonstrate that cancer interception and prevention strategies will need to be tailored to individual patients.

- **What was the impact on other disciplines?**

Nothing to report.

- **What was the impact on technology transfer?**

Nothing to report.

- **What was the impact on society beyond science and technology?**

Nothing to report.

5. Changes/Problems

- **Changes in approach and reasons for change**

No changes were made to the original research plan.

- **Actual or anticipated problems or delays and actions or plans to resolve them**

No problems have been encountered.

- **Changes that had a significant impact on expenditures**

Nothing to report.

- **Significant changes in use or care of human subjects, vertebrate animals, biohazards, and/or select agents**

Nothing to report.

- **Significant changes in use or care of human subjects**

Nothing to report.

- **Significant changes in use or care of vertebrate animals**

Nothing to report.

- **Significant changes in use of biohazards and/or select agents**

Nothing to report.

6. Products

- **Publications, conference papers, and presentations**
  - **Journal publications.**

Nothing to report.

  - **Books or other non-periodical, one-time publications.**

Krysan K, Tran LM, Grimes BS, Walser TW, Wallace WD, Dubinett SM. Evaluation of progression-associated neoepitopes and immune contexture in pulmonary premalignancy. *AACR Annual Meeting 2017, Washington, DC (oral presentation). Abstract number: 17-A-7256.*

  - **Other publications, conference papers, and presentations.**

Nothing to report.

- **Website(s) or other Internet site(s)**

Nothing to report.

- **Technologies or techniques**

Nothing to report.

- **Inventions, patent applications, and/or licenses**

Nothing to report.

- **Other Products**

Nothing to report.

7. Participants & Other Collaborating Organizations

- **What individuals have worked on the project?**

Kostyantyn Krysan.
No change.
Linh M. Tran.
No change.
William D. Wallace.
No change.

- **Has there been a change in the active other support of the PD/PI(s) or senior/key personnel since the last reporting period?**

Nothing to report.

- **What other organizations were involved as partners?**

Nothing to report.

8. Special Reporting Requirements

Nothing to report.

9. Appendices

None.